THE ANALYST.

VOL. V. MAY, 1878. No. 3.

ON REPEATED ADJUSTMENTS, AND ON SIGNS OF RESIDUALS.

BY E. L. DE FOREST, M. A., WATERTOWN, CONN.

LET any consecutive terms in a series be denoted by

$$\ldots\ldots u_{-2},\ u_{-1},\ u_0,\ u_1,\ u_2,\ \ldots\ \&c.$$

Suppose them to be adjusted by the formula

$$u'_0 = l_0u_0+l_1(u_1+u_{-1})+l_2(u_2+u_{-2})+l_3(u_3+u_{-3})+l_4(u_4+u_{-4}), \quad (1)$$

and let the adjusted terms be again adjusted by the formula

$$u'_0 = L_0u_0+L_1(u_1+u_{-1})+L_2(u_2+u_{-2}). \quad (2)$$

By the first operation, five consecutive adjusted terms will be

$$\begin{aligned}
u'_{-2} &= l_0u_{-2}+l_1(u_{-1}+u_{-3})+l_2(u_0+u_{-4})+l_3(u_1+u_{-5})+l_4(u_2+u_{-6})\\
u'_{-1} &= l_0u_{-1}+l_1(u_0+u_{-2})+l_2(u_1+u_{-3})+l_3(u_2+u_{-4})+l_4(u_3+u_{-5})\\
u'_0 &= l_0u_0+l_1(u_1+u_{-1})+l_2(u_2+u_{-2})+l_3(u_3+u_{-3})+l_4(u_4+u_{-4})\\
u'_1 &= l_0u_1+l_1(u_2+u_0)+l_2(u_3+u_{-1})+l_3(u_4+u_{-2})+l_4(u_5+u_{-3})\\
u'_2 &= l_0u_2+l_1(u_3+u_1)+l_2(u_4+u_0)+l_3(u_5+u_{-1})+l_4(u_6+u_{-2}),
\end{aligned}$$

and by the second operation, a readjusted term will be

$$u''_0 = L_0u'_0+L_1(u'_1+u'_{-1})+L_2(u'_2+u'_{-2}).$$

Substituting in this the values of u'_0, u'_1, &c., as above, we get

$$\left.\begin{aligned}
u''_0 = {} & [L_0l_0+L_1(l_1+l_1)+L_2(l_2+l_2)]u_0\\
&+[L_0l_1+L_1(l_0+l_2)+L_2(l_1+l_3)](u_1+u_{-1})\\
&+[L_0l_2+L_1(l_1+l_3)+L_2(l_0+l_4)](u_2+u_{-2})\\
&+[L_0l_3+L_1(l_2+l_4)+L_2l_1](u_3+u_{-3})\\
&+(L_0l_4+L_1l_3+L_2l_2)(u_4+u_{-4})\\
&+(L_1l_4+L_2l_3)(u_5+u_{-5})+L_2l_4(u_6+u_{-6}).
\end{aligned}\right\} \quad (3)$$

It thus appears that the re-adjusted term u''_0 is a linear function of the original terms u_0, u_1, &c., and that the expression (3) may itself be regarded as an adjustment formula which replaces and is equivalent to the two formulas (1) and (2) used successively. Furthermore, if the order of procedure

is reversed, and the original series is adjusted, first by (2), and this again by (1), it will be found that the re-adjusted term is given by precisely the same formula (3). The operations of successive adjustment, then, are of such a kind that the order of operation is immaterial. This has already been remarked by me in the *Smithsonian Report* of 1873, p. 337, though coupled with an erroneous inference about the probable error of such an adjustment, for correction of which see ANALYST for May 1877, p. 84. The fact that (3) replaces (1) and (2) without regard to the order of their application is one which holds good for any values we please to assign to the coefficients l_0, l_1, &c., and L_0, L_1, &c., and also holds good without regard to the nature of the series u_0, u_1, &c. The coefficients in (3) are functions of those in (1) and (2), and may be computed when these are given. If (1) and (2) are adjustment formulas each of which gives an exact result when applied to a series of equidistant terms of an order not higher than the third, that is, a series whose third differences are constant, then the terms of such a series adjusted by (1) will be the same as the original terms, and these re-adjusted by (2) will still remain the same, so that (3), which is equivalent to (1) and (2) together, will be also an adjustment formula which gives exact results when applied to such series. Its coefficients, therefore, will be such as to satisfy the two relations (2) given at p. 80 of the ANALYST referred to, these relations being satisfied, in the case supposed, by the coefficients of the component formulas (1) and (2).

Results similar to the above will be obtained when formulas other than (1) and (2), that is, formulas including some other numbers of terms, are successively used. We may generalize them as follows. Let the two component formulas be

$$u'_0 = l_0 u_0 + l_1 (u_1 + u_{-1}) + l_2 (u_2 + u_{-2}) + \ldots\ldots + l_m (u_m + u_{-m}) \quad (4)$$

$$u'_0 = L_0 u_0 + L_1 (u_1 + u_{-1}) + L_2 (u_2 + u_{-2}) + \ldots\ldots + L_n (u_n + u_{-n}), \quad (5)$$

and for convenience let the second one be the shorter, or $n <$ or $= m$. A comparison of (3) shows that the resultant formula will be

$$\left.\begin{array}{l} u''_0 = [L_0 l_0 + L_1(l_1 + l_1) + L_2(l_2 + l_2) + \ldots\ldots + L_n(l_n + l_n)](u_0 \\ \quad + [L_0 l_1 + L_1(l_0 + l_2) + L_2(l_1 + l_3) + \ldots\ldots + L_n(l_{n-1} + l_{n+1})](u_1 + u_{-1}) \\ \quad + [L_0 l_2 + L_1(l_1 + l_3) + L_2(l_0 + l_4) + \ldots\ldots + L_n(l_{n-2} + l_{n+2})](u_2 + u_{-2}) \\ \quad + [L_0 l_3 + L_1(l_2 + l_4) + L_2(l_1 + l_5) + \ldots\ldots + L_n(l_{n-3} + l_{n+3})](u_3 + u_{-3}) \\ \quad \cdot\ \cdot \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad + L_n l_{-m}(u_{m+n} + u_{-(m+n)}) \end{array}\right\} \quad \ldots (6)$$

The mode of construction of this expression is easily seen. The coefficients l are arranged in columns from the top downward, in the order in which they stand in the series, $l_m, l_{m-1}, l_{m-2}, \ldots l_3, l_2, l_1, l_0, l_1, l_2, l_3, \ldots l_{m-1}, l_m$.

It must be remembered that all values of l beyond l_m are zero, and also that wherever the expression gives to l a negative index figure, it must be changed to positive, since $l_{-i} = l_i$.

In this way, by taking $m = 4$ and $n = 2$, we find that (6) reduces to (3).

If a series is adjusted twice by the same formula, then (5) is identical with (4), and (6) is somewhat simplified, because we now have

$$L_i l_k = L_k l_i,$$

so that (6) may be written as follows.

$$\left.\begin{aligned}u''_0 = {} & [l_0 l_0 + 2(l_1 l_1 + l_2 l_2 + l_3 l_3 + l_4 l_4 + \dots + l_m \ l_m)]u_0 \\ & \quad + 2(l_0 l_1 + l_1 l_2 + l_2 l_3 + l_3 l_4 + \dots + l_{m-1} l_m)(u_1 + u_{-1}) \\ & + [l_1 l_1 + 2(l_0 l_2 + l_1 l_3 + l_2 l_4 + l_3 l_5 + \dots + l_{m-2} l_m)(u_2 + u_{-2}) \\ & \quad + 2(l_1 l_2 + l_0 l_3 + l_1 l_4 + l_2 l_5 + \dots + l_{m-3} l_m)(u_3 + u_{-3}) \\ & + [l_2 l_2 + 2(l_1 l_3 + l_0 l_4 + l_1 l_5 + l_2 l_6 + \dots + l_{m-4} l_m)(u_4 + u_{-4}) \\ & \quad + 2(l_2 l_3 + l_1 l_4 + l_0 l_5 + l_1 l_6 + \dots + l_{m-5} l_m)(u_5 + u_{-5}) \\ & \quad \dots\dots\dots\dots\dots\dots\dots\dots \\ & \quad + 2 l_{m-1} l_m (u_{2m-1} + u_{-(2m-1)}) + l_m l_m (u_{2m} + u_{-2m}).\end{aligned}\right\} \quad (7)$$

In the successive products ll in the horizontal lines of this formula, the sequence of the indices of the first l, and also of those of the second l, is easily seen. Let us apply this to a simple case. The 5-term formula of Table A. (ANALYST, May 1877, p. 81), in exact numbers is

$$u'_0 = \tfrac{1}{35}[17u_0 + 12(u_1 + u_{-1}) — 3(u_2 + u_{-2})] \quad (8)$$

so that the coefficients l, in decimals of five places are

$$l_0 = .48572 \qquad l_1 = .34286 \qquad l_2 = —.08572.$$

Forming the products

$$l_0 l_0, \quad l_0 l_1, \quad l_0 l_2, \quad l_1 l_1, \quad l_1 l_2, \quad l_2 l_2,$$

and putting $m = 2$, we get by formula (7)

$$\begin{aligned}u''_0 = {} & .48572u_0 + .27430(u_1 + u_{-1}) + .03427(u_2 + u_{-2}) \\ & —.05878(u_3 + u_{-3}) + .00735(u_4 + u_{-4}). \quad (9)\end{aligned}$$

This is an adjustment formula such that if a given series is adjusted by it once, the result is the same as though it had been adjusted by (8) twice. Again, if we suppose a series to be adjusted by both (8) and (9) successively, the equivalent single formula is found by designating their coefficients by L and l respectively, forming all the products of every L by every l, and putting $m = 4$ and $n = 2$. Formula (6) then gives

$$\begin{aligned}u'''_0 = {} & .41814u_0 + .29305(u_1 + u_{-1}) + .04827(u_2 + u_{-2}) — .03780(u_3 + u_{-3}) \\ & —.01952(u_4 + u_{-4}) + .00756(u_5 + u_{-5}) — .00063(u_6 + u_{-6}), \\ & \dots \quad (10)\end{aligned}$$

and this result is denoted by u'''_0 because one adjustment of a series by this formula is equivalent to three successive adjustments by the original formula (8). In the same way we may combine (10) with (8), (9) or (10), and thus get the equivalents of four, five and six applications of (8). The process

has indeed been extended much farther. It is not so difficult or tedious as might be supposed, for it has been found that the coefficients of the extreme terms of the formula soon become so small as to disappear, that is, they do not affect the first five places of decimals. Thus in (10) the coefficient of u_6 is only seen in the fourth place. At the successive steps of the process the correctness of the work is readily tested by the condition that for every formula obtained we must have

$$1_0+2(1_1+1_2+1_3+\&c.)=1,$$

within such narrow limits of error as are to be expected from the use of logarithms in forming the products. And at each step, before proceeding farther, it is best to make the formula satisfy this condition exactly, by altering one, two or three of the coefficients, if necessary, to the extent of a single unit in the fifth decimal place. The various products have been carried to five, and sometimes to six, places of decimals. The resulting coefficients are shown in the accompanying table I., to three places only. The numbers of successive applications of the original formula (8) are found in the upper line of the table, and the corresponding coefficients are in the columns

TABLE I.

	1	2	3	4	5	6	8	12	16	32
l_0	.486	.486	.418	.396	.372	.354	.330	.296	.274	.230
l_1	.343	.274	.293	.280	.277	.271	.261	.246	.234	.205
l_2	—.086	.034	.048	.077	.090	.102	.116	.131	.139	.145
l_3		—.059	—.038	—.034	-.023	-.014	.002	.026	.042	.072
l_4		.008	—.019	—.024	-.030	-.032	-.031	-.024	-.015	.014
l_5			.008	.000	-.004	-.009	-.017	-.024	-.026	-.017
l_6			—.001	.004	.004	.003	.000	-.008	-.014	-.022
l_7				—.001	.001	·002	.003	.002	-.002	-.014
l_8					-.001	-.000	.001	.003	.003	-.004
l_9									.002	.002
l_{10}										.003
l_{11}										.001
r	.697	.629	.596	.574	.558	.544	.524	.497	.479	.438
ρ	.297	.102		.0513		.0331		.0160		.0058

below. The coefficients in the last column, for instance, are those of an adjustment formula such that if a given series is adjusted by it once, the result is the same as if it had been adjusted by the original formula (8) thirty-two times in succession, or by (9) sixteen times. This 32-fold formula, if given exactly, would include no less than 129 terms, with 65 different coefficients, but only 12 coefficients are large enough to appear in the third decimal place, and only four more would affect the fourth place. At the bottom of

the table are given the ratios of error and of irregularity

$$r=\frac{\varepsilon'}{\varepsilon},\qquad\qquad \rho=\frac{(\Delta'_4)}{(\Delta_4)}.$$

The ratio r diminishes slowly as the number of repetitions is increased, while ρ diminishes much faster. For any large number of repetitions, the form of the series of values of l approximates somewhat to that of the formulas of Table B. (ANALYST, p. 82.) For instance, the large coefficients of the 12-fold and 32-fold formulas of Table I. approach very near to those of the 11-term and 15-term formulas of Table B., and their ratios r and ρ are also nearly alike.

A similar process of repetitions has been applied to the 5-term formula of Table B., which in its exact form is

$$u'_0=\tfrac{1}{195}[111u_0+56(u_1+u_{-1})—14(u_2+u_{-2})]. \tag{11}$$

(*Smithsonian Report* of 1871, p. 333.) The resultant coefficients are given in the first half of Table II. The second half contains similar results of repetitions of the formula, found at p. 289 of the *Report*,

$$u'_0=\tfrac{1}{10}[4(u_0+u_1+u_{-1})—(u_2+u_{-2})]. \tag{12}$$

TABLE II.

	1	2	4	6	12	1	2	4	6	12
l_0	.570	.500	.416	.374	.310	.400	.500	.392	.344	.284
l_1	.287	.285	.285	.276	.253	.400	.240	.264	.261	.239
l_2	-.072	.001	.059	.089	.125	-.100	.080	.101	.115	.136
l_3		-.041	-.038	—.023	.015		-.080	-.037	—.008	.035
l_4		.005	-.019	—.029	—.028		.010	-.022	—.030	—.019
l_5			.003	—.004	—.021			-.008	—.016	—.026
l_6			.002	.003	—.004			.008	.004	—.011
l_7				.001	.003			-.002	.002	.000
l_8					.002				.001	.003
l_9									—.001	.001
r	.707	.645	.588	.558	.508	.707	.625	.563	.532	.486
ρ	.229	.122	.0622	.0411	.0195	.446	.239	.0829	.0365	.0134

It will be seen that although the coefficients l of the two original formulas (11) and (12) differ from each other very considerably, yet after 12 applications they approach pretty nearly to each other and also to those of the 12-fold formula of Table I. The idea is thus suggested, that when any 5-term formulas similar to (8), (11) and (12) are applied a large number of

times in succession, the resultant coefficients approximate to a common form, and that they resemble the coefficients of Table B. a good deal more closely than those of Tables A. or C. This may perhaps be regarded as an argument in favor of the use of the formulas of Table B. for purposes of adjustment. It is reinforced by the fact that when other formulas are repeatedly used, they take a similar outline. The 9-term formula of Table A., in exact numbers, is

$$u'_0=\tfrac{1}{231}[59u_0+54(u_1+u_{-1})+39(u_2+u_{-2})+14(u_3+u_{-3})-21(u_4+u_{-4})],$$

and if it is used four times in succession, the coefficients of the equivalent formulas are, to three places of decimals,

.202	.185	.141	.086	.035	—.001	—.018
	—.018	—.011	—.004	.001	.002	.001.

These differ but little from the coefficients in the 17-term formula of Table B. Again, the 13-term formula of Table A., in exact numbers, is

$$u'_0=\tfrac{1}{143}[25u_0+24(u_1+u_{-1})+21(u_2+u_{-2})+16(u_3+u_{-3})$$
$$+\ 9(u_4+u_{-4})+\ 0(u_5+u_{-5})-11(u_6+u_{-6})],$$

and, after four successive applications of it, the resultant formula is found to have the following coefficients:

.138	.132	.117	.095	.069	.043	.021
	.004	—.008	—.013	—.013	—.010	—.006
	—.003	—.001	.001	.001	.001	.001.

There is no very great difference between these and the coefficients of the 25-term formula of Table B.

A comparison of the values of r and ρ in Tables I. and II. affords some rather unexpected results. Although the original formula (12) gives values to r and ρ as large as, or larger than, either of the original formulas (8) and (11) do, yet after twelve successive applications, it gives smaller values than the others do. In other words, formula (8), which was found by the method of least squares, and makes, when the conditions assumed are fulfilled, a more accurate single adjustment than any other 5-term formula, is not the one which gives the greatest accuracy when the adjustment is repeated several times, or even when it is repeated once only. Likewise formula (11), which under the same conditions makes the smoothest single adjustment of any, is not the one which gives greatest smoothness when the adjustment is repeated.

Other calculations have been made for the purpose of learning what the results will be when the 25-term formulas of Tables A., B. and C. are used in succession, that is, when any one of them is used twice, or when any two are used one after the other. Each of the complete resultant formulas includes 49 terms, with 25 different coefficients, and from 17 to 25 of these

are large enough to affect the third decimal figure. It is thought unnecessary to give any of them here in full, but the ratios r and ρ for all of them are shown in Table III. For instance, when the 25-term formulas of Tables A. and B. are used in succession, no matter which is used first, the error-ratio and irregularity ratio of the resulting adjustment will be

$$r = .281 \qquad\qquad \rho = .000191.$$

TABLE III.

	Values of r.			Values of ρ.		
	A	B	C	A	B	C
A	.266			.002649		
B	.281	.313		.000191	.000324	
C	.272	.291	.279	.000301	.000185	.000313

When the 25-term formula of Table C. is applied twice, the ratios for the result will be $r = .279$ $\qquad \rho = .000313$,
and so on. As all the values of r in Table III. are larger than .250, it will obviously be difficult, if not impossible, ever to adjust a series by these methods so as to reduce the error-ratio below $\frac{1}{4}$. It diminishes but very slowly with each repetition, the ratio for the second application of A., for instance, being .266, while for the first application Table A. shows it to be .300, and we have

$$\frac{.266}{.300} = .887,$$

so that the probable error of the second adjustment is little less than nine tenths that of the first one.

We also see that the smallest value of ρ in Table III. is not, as might have been expected, that which results from two applications of the B formula. Smaller values are given not only by two applications of the C, but also by either A and B, A and C, or B and C.

SIGNS OF RESIDUALS.—In the ANALYST for Jan. 1878, p. 3, it was shown by me that when a series has been adjusted by a formula whose error-ratio r is small, a safe limit being $r < \frac{1}{2}$, the probability that the signs of any two adjacent residual errors will be alike is approximately

$$q = \frac{1-2r^2}{2-2r^2}, \tag{13}$$

and as a consequence of this, that the most probable number M of signs falling in groups of only one or two signs each, in a periodic series, is given by

$$\left.\begin{array}{c} p_1 = (1-q)^2, \qquad\qquad p_2 = p_1 q, \\ M = (p_1 + 2p_2)N \pm .6745\sqrt{[p_1(1-p_1) + 4p_2(1-p_2)]}\sqrt{N}, \end{array}\right\} \tag{14}$$

where N is the whole number of terms in the series. The probable number falling in groups of more than two is

$$M' = N - M.$$

The above will answer also for non-periodic series, if $N-\frac{5}{2}$ is put in place of N, in which case the first and last signs of the series, and as many adjacent ones as happen to be like them, are not counted. If we write

$$M' = aN \pm b\sqrt{N},$$

the values of a and b, for different values of r, will be as here given in Table IV.

TABLE IV.

r	a	b	a_1	b_1
.25	.450	.55	.467	.34
.30	.426	.56	.451	.34
.35	.396	.56	.430	.33
.40	.359	.57	.405	.33
.45	.314	.58	.373	.33
.50	.259	.58	.333	.32

The process by which (14) was obtained assumes that any four consecutive signs may be regarded as independent of each other, except as adjacent ones are connected by the relation (13). This is only approximately true. For the purpose of testing an adjustment, it may be better, and is simpler, to consider only the number of permanences of sign, that is, the number of cases where adjacent signs are alike, in passing along the series of residuals from one end to the other. The most probable number of permanences will be, in a periodic series,

$$P = qN \pm .6745\sqrt{[q(1-q)N]}, \qquad (15)$$

and the same for a non-periodic series, only putting $N-1$ in the place of N.

If (15) is written

$$P = a_1 N \pm b_1 \sqrt{N}$$

the values of a_1 and b_1, for different values of r, will be found in Table IV. For example, in the observed series of signs at p. 5 of the ANALYST referred to, we have $r = .372$, $N = 70$ and by interpolation in Table IV., $a = .380$, $b = .57$, $a_1 = .419$, $b_1 = .33$, so that computation gives

$$M' = .380 \times 67.5 \pm .56\sqrt{(67.5)} = 25.7 \pm 4.6,$$
$$P = .419 \times 69 \pm .33\sqrt{(69)} = 28.9 \pm 2.7.$$

Observation by actual count in the series shows

$$M' = 22, \qquad P = 25.$$

Both fall short of the computed numbers, but the deviation is on the safe side, and must be considered accidental. It is only when the observed values of M' and P exceed the computed values, by more than the probable errors, that we are to infer the existence of a danger that the adjusted series has been smoothed out too much, so as to deviate materially from its true law.